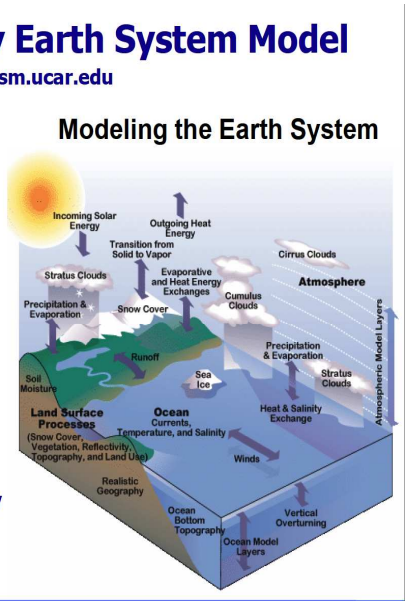
SDCI Net: An integrated study of datacenter networking

John Dennis
dennis@ucar.edu

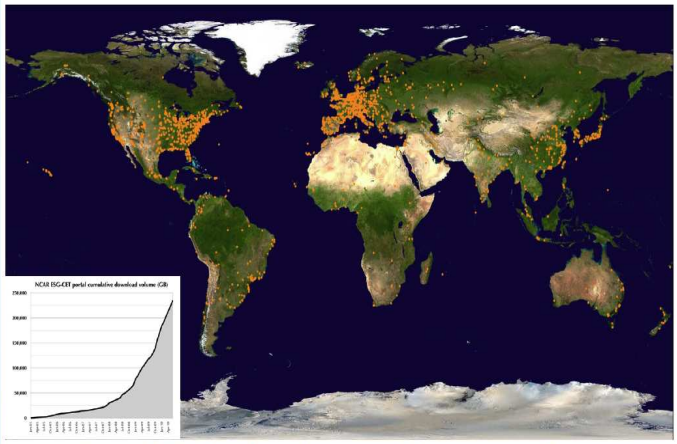
The Community Earth System Model

www.cesm.ucar.edu

- CESM: a set of different geophysical component models that exchange boundary data via a coupler
- Code base developed over 20+ yrs: runs on multiple platforms, resolutions and model configurations
- CESM is used to:
  - Explore Earth climate history and processes responsible for variability and change
  - Estimate future of environment for policy formulation
- Developed by NCAR NSF, DOE, Universities, National Laboratories
- Fully documented, frequently and freely distributed, fully supported releases
- Capacity Building (e.g., tutorials and workshops)



A Community Resource



Over 3,000 sites from 130+ countries

>230 Tb since January 2005

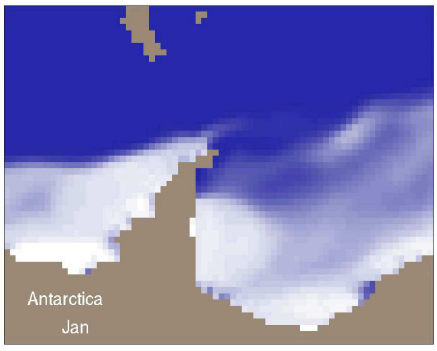
Courtesy Gary Strand

How to efficient utilize very large computers?

NSF PetaApps project (OCI-079206)

- PetaApps project: Interactive Ensemble
  - Kinter, Stan (COLA)
  - Kirtman (U of Miami)
  - Collins, Yelick (Berkley)
  - Bryan, Dennis, Loft, Vertenstein (NCAR)
  - Bitz (U of Washington)
- Ultra-High resolution climate
  - Explore impact of weather noise on Climate
  - Explore technical/Computer Science issues
- ~99,000 core Cray XT5 system at NICS [Kraken]
- Large TG allocation:
  - 35M CPU hours [2009]
  - 28M CPU hours [2011]

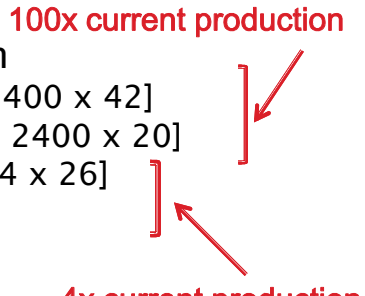
Weddell Sea: Ice thickness (1°)



The Role of Climate System Noise in Climate Simulations

Large scale PetaApps run

- Component configuration
  - 0.1° Ocean model [3600 x 2400 x 42]
  - 0.1° Sea-ice model [3600 x 2400 x 20]
  - 0.5° Atmosphere [576 x 384 x 26]
  - 0.5° Land [576 x 384]
- Statistics
  - ~18M CPU hours
  - 5844 cores for 4-5 months
  - ~100 TB of data generated
  - 0.5 to 1 TB per wall clock day generated



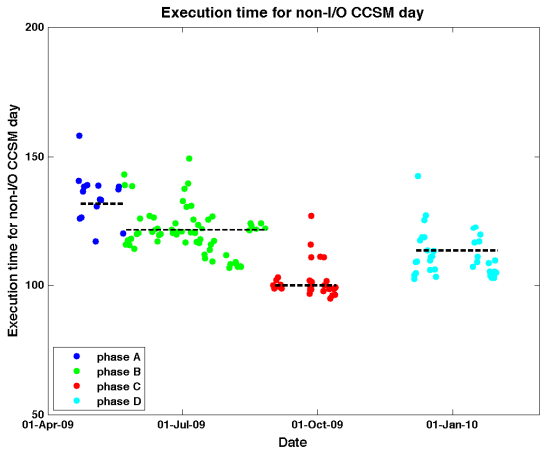
Large scale PetaApps run (con't)

- Work flow
  - Run on Kraken (NICS)
  - Transfer output from NICS to NCAR (100 – 180 MB/sec sustained)
  - Archive on HPSS
  - Data analysis using 55 TB project space at NCAR

Issues/challenges with runs

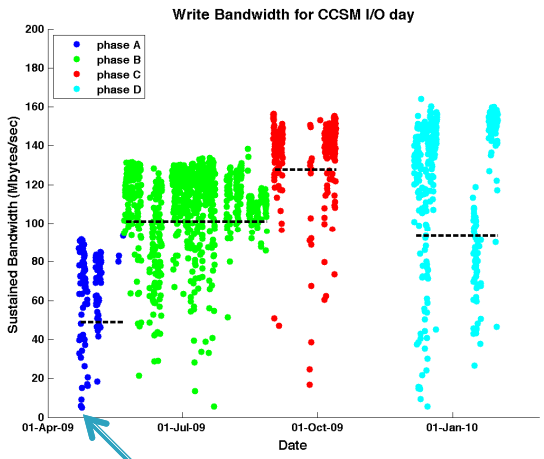
- Very large variability with I/O performance
  - 2-10x slowdown common
  - 300x slowdown was observed
  - Interference with other jobs?

Execution time for non-I/O CESM day



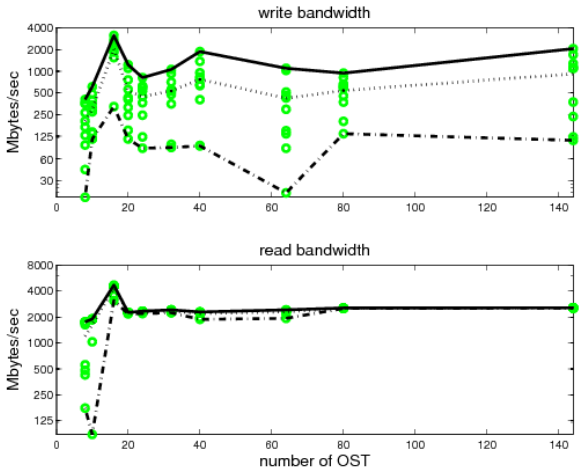
Network contention/congestion in message passing network?

Write bandwidth for CESM I/O day



What happened on April 26, 2009 ?

Parallel I/O variability on Kraken



Potential climate experiments

- Granted:
  - PetaApps: TRAC-2, 28M CPU hours, ~120 TB, NICS, [2011-2012]
  - PRAC, 590M CPU hours, ~6,000 TB, NCSA, [2012-2013]
- Under Review:
  - SciDAC, 7M CPU hours, ~51 TB, NERSC [2012]
- Planned:
  - Breakthrough Science, 45M CPU hours, 200 TB, NWSC [2012]

Potential climate experiments (con't)

- Sources of Data
  - NICS
  - NERSC
  - NWSC (NCAR-Wyoming Supercomputer Center)
  - NCSA
- Sinks of Data
  - NCAR
  - NWSC
  - NCSA
- Consumers of Data
  - NCAR
  - Colorado State University
  - University of California Berkley
  - George Mason University
  - University of Miami
  - University of Washington
  - Texas A&M
  - University of Hawaii
  - Woods Hole Oceanographic Institute

Timeline for 1st-year

- Configure gridFTP logging for data collection [Fall,Winter 2011]
- Participate in 100Gb ANI testbed work [Dec 2011]
- Initial profiling of MPI applications [Jan 2012]
- Through analysis of MPI application profiling [March – July 2012]
- Collection of gridFTP usage data [Jan – Sept 2012]

Questions?

dennis@ucar.edu

High-Resolution CCSM/CESM

- LLNL Grand Challenge project
  - 0.25° ATM, LND + 0.1° OCN, ICE
  - Participants:
    - Bader [PI] , McClean, Ivanova, Boyle (LLNL)
    - Bryan, Dennis, Vertenstein, Craig, Norton (NCAR)
    - Jones (LANL)
    - Worley (ORNL)
    - Jacob (ANL)
  - 20 year run
- NSF PetaApps project: Interactive Ensemble
  - 0.5° ATM, LND + 0.1° OCN, ICE
  - Participants:
    - Kinter [PI], Stan (COLA)
    - Kirtman (U of Miami)
    - Collins, Yelick (Berkley)
    - Bryan, Dennis, Loft, Vertenstein (NCAR)
    - Bitz (U of Washington)
  - Teragrid Allocations
    - 35M SU (2009)
    - 28M SU (2011)
  - 155 year control run + several 50 year branches

IPCC & AR5

- Intergovernmental Panel on Climate Change (IPCC)

  "The IPCC assesses the scientific, technical and socio-economical information relevant for the understanding of the risk of human-induced climate change"
- The Fifth Assessment Report (AR5)
- CESM plans
  - Control: 1 run x 1300 years
  - Historical: 40 runs x 55 years
  - Future: 180 runs x 95 years

  20,600 years of simulation/ 564 TB

Climate Modeling?

- Coupled models Atmosphere-Ocean-Sea Ice-Land
- Challenges of climate:
  - Need many years -->
  - Limits resolution -->
  - Limits parallelism
- Community Earth System Model (CESM)
  Formally: Community Climate System Model (CCSM)

Acknowledgements

- NCAR:
  - D. Bailey
  - F. Bryan
  - T. Craig
  - B. Eaton
  - J. Edwards [IBM]
  - N. Hearn
  - K. Lindsay
  - N. Norton
  - M. Vertenstein
- COLA:
  - J. Kinter
  - C. Stan
- U. Miami
  - B. Kirtman
- U.C. Berkeley
  - W. Collins
  - K. Yelick (NERSC)
- U. Washington
  - C. Bitz
- NICS:
  - M. Fahey
  - P. Kovatch
- ANL:
  - R. Jacob
  - R. Loy
- LANL:
  - E. Hunke
  - P. Jones
  - M. Maltrud
- LLNL
  - D. Bader
  - D. Ivanova
  - J. McClean (Scripps)
  - A. Mirin
- ORNL:
  - P. Worley

and many more…

- Grant Support:
  - DOE
    - DE-FC03-97ER62402 [SciDAC]
    - DE-PS02-07ER07-06 [SciDAC]
  - NSF
    - Cooperative Grant NSF01
    - OCI-0749206 [PetaApps]
    - CNS-0421498
    - CNS-0420873
    - CNS-0420985
- Computer Allocations:
  - TeraGrid TRAC @ NICS
  - DOE INCITE @ NERSC
  - LLNL Grand Challenge
- Thanks for Assistance:
  - Cray, NICS, and NERSC